Solving Missing Data Problems in Emissions Modeling

Crosswalk Labs is using generative AI to fill gaps in building property data for emissions modeling. Intelligent agents enrich and validate attributes like square footage, enabling more accurate and scalable climate analytics.

Filling Data Gaps with AI Agents

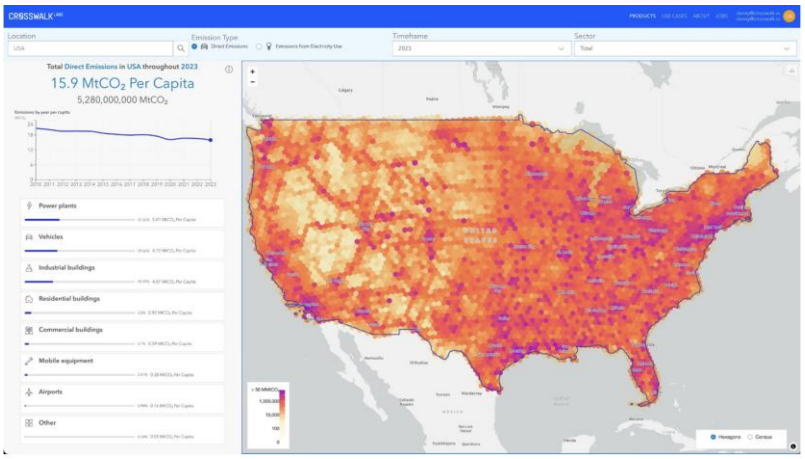
Deploys serverless GenAI agents (Amazon Bedrock + Lambda/APIGateway) to enrich, validate, and correct building attributes like square footage.

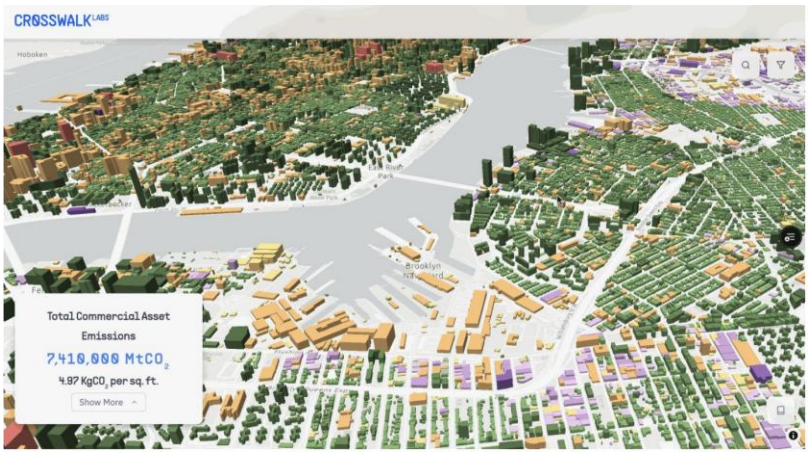
Improving Emissions Modeling

Produces scalable, high-quality building-level GHG estimates for climate modeling, climate reporting, and real-estate analytics.

Key Features

- **AI-Driven Data Enrichment:** Uses GenAI models and web search APIs to fill gaps and correct outliers in building datasets.
- **Scalable and Automated:** Serverless infrastructure ensures low-latency, on-demand performance across large national datasets.
- **Open and Transparent:** Supports open data standards and supports data available on open.crosswalk.io.





Applications

- **Climate Compliance:** Generates more accurate emissions data for sustainability reporting and disclosure.
- **Urban Planning and Policy:** Supports city and national climate strategies with reliable building-level GHG data.
- **Real Estate Analytics**: Powers commercial tools with enriched building datasets for investment and asset management.

Team:
Crosswalk Labs (Daniel M. Sheehan, Jason Burnett, Dr. Phil DeCola, Dr. Victoria Hunt, Dr. Geoff Roest, Dr. Anastasia Montgomery)

Technical Challenges

A primary technical challenge that the team encountered with this work was that finding and correcting errors in large datasets of building-level attributes is time consuming and difficult, especially when a manual approach is attempted. When error-handling is inadequate, this contributes to inaccuracies in final modeled estimates of Greenhouse Gas Emissions (GHGs) and climate analytics.

Observations and Lessons Learned

This project reaffirmed the potential of combining generative AI with geospatial datasets, especially for modelling applications where data completeness and accuracy is very important.

Next Steps

The Crosswalk team plans to benchmark against manually validated records and continue fine-tuning the method for error-handling. Further, they believe that this type of solution is not only applicable to square footage, but could also be used for building type, activity data, etc. They plan on evaluating the completeness and accuracy across multiple datasets to cross-check their work. For example, they will evaluate across open datasets such as OpenStreetMap and Overture Maps data vs. National Structure Inventory data available from the Federal Emergency Management Agency (FEMA).

Customers

This work supports Crosswalk Labs' emissions estimation platform, which is used by local governments, advocacy groups, academic researchers and the private sector. Enhanced building attribute coverage and accuracy increases the fidelity of their emissions estimates, enabling more accurate reporting.